reprinted from:
A. Plückthun, Studying protein structure and function by directed evolution. In: Protein Dynamics, Function and Design, O. Jardetzky and J. F. Lefèvre, Eds. (Plenum Press, New York, 1998), vol. 301, pp. 37-57.

STUDYING PROTEIN STRUCTURE AND FUNCTION BY DIRECTED EVOLUTION

Examples with Engineered Antibodies

Andreas Plückthun*

Biochemisches Institut
Universität Zürich
Winterthurerstr. 190, CH-8057 Zürich, Switzerland

ABSTRACT

Directed molecular evolution is a powerful strategy for investigating the structure and function of proteins. When a function, such as ligand binding, can only be carried out by the native state of the protein, the biological selection for this function can be used to improve structural properties of the protein. Thus, thermodynamic stability and folding efficiency, which is the ability to avoid aggregation during folding, can be optimized. Three methods of selection are reviewed: phage display, selectively infective phages (SIP) and ribosome display, a cell-free method. Examples for optimizing antibody stability are discussed. In one case, antibodies have been generated under evolutionary pressure, which are stable in the absence of any disulfide bond, in the other case, a kink in the first strand of the beta-sandwich of kappa domains has been optimized.

1. INTRODUCTION

The use of molecular evolution technologies is starting to provide a new perspective in the study of the structure, stability and function of proteins. Representative examples in several application areas such as enzyme activity or antibody affinity illustrate the power of this approach (Hall and Knowles, 1976; Stemmer, 1994a; Low et al., 1996; Baca et al., 1997; Hanes and Plückthun, 1997; Moore et al., 1997). Like the natural evolution of pro-

* Tel: (+41-1) 635 5570; FAX: (+41-1) 635 5712; E-mail: plueckthun@biocfebs.unizh.ch

teins, directed evolution alternates between creation of diversity and selection. Depending on the problem at hand, diversity may be focused on certain regions in a protein, or the whole sequence may be sprinkled with mutations.

Out of this diversity, superior molecules need to be selected. In most non-enzyme cases, the only directly selectable function is binding to a ligand, but many structural features which affect the amount of the protein molecule produced or its stability indirectly affect selection efficiency, as long as ligand binding is limited to the correctly folded molecule. It is this coupling of function to the native state which has been exploited to optimize protein structure in the examples reviewed here.

While the evolutionary approach is only one of several genetic approaches for studying protein structure and function, it has provided a new line of attack for the elucidation of effects which are based on exquisitely complex interactions between amino acids, each contributing only small energies to stability, ligand binding or transition state stabilization. Thus, the technology of directed molecular evolution provides a new sampling mechanism of variants along a pathway of improved function.

In the past, the combination of site directed mutagenesis and biophysical investigations has provided enormous insight into the architecture and design of proteins, by allowing the investigator to change specific amino acids and testing their influence on the parameter of interest, be it stability, folding kinetics, or functional properties of the protein. While very successful for testing hypotheses of the function of a small number of residues, it is only efficient as long as a very good theoretical understanding of the problem under investigation is already available, such that relevant mutants can be planned. Examples where site directed mutagenesis has been instrumental is in the study of enzyme mechanisms, where from a knowledge of the structure those amino acids can often be pinpointed which might be involved in the mechanism (Fersht and Wells, 1991). An important extension is the use of multiple mutants to elucidate synergistic effects—the so-called double mutant cycles (Horovitz, 1996). In this strategy, the relative contribution of one mutant in the presence or absence of another is quantitatively compared, and the relative interaction energy of the two amino acids can be elucidated.

A very systematic mutagenesis of every single residue and multiple combinations thereof can be carried out to define e.g. extended binding sites. This has frequently been called "alanine scanning", since alanine is the ideal residue to which to change a particular residue, alanine being compatible with most secondary structures, sufficiently hydrophilic and not carrying a functional group in the side chain (Ward et al., 1990; Wells et al., 1993). This strategy has shown great success in e.g. defining hormone-receptor interactions or antibody-antigen interactions (Jin et al., 1992; Wells et al., 1993).

The use of such directed mutagenesis approaches has also been successful in the study of protein stability and the temporal events of protein folding (Fersht, 1995; 1997). While very informative, these experiments are quite labor intensive and are not very efficient in uncovering new relations in protein stability and internal interactions, unless carried out on a very large scale.

The main problem in all site-directed mutagenesis strategies is the combinatorial explosion of mutants one might be tempted to make. Consequently, at the present level of understanding, the rational "improvement" of ligand binding or protein stability by cycles of directed mutagenesis, their analysis and deduction of further mutations has not been very successful, and this strategy of the "protein design cycle" has the unfortunate property of being both labor intensive and slow. It is this problem that the technology of "directed evolution" may help to solve, by providing a highly efficient sampling pathway of informative mutants.

2. THE CHALLENGES IN DIRECTED MOLECULAR EVOLUTION

In the strategy of directed molecular evolution nature is imitated by using a succession of mutagenesis and selection. Thus, there are three technical challenges to overcome.

The first is to generate a sufficient number of diverse mutants. While, over the last few years, a number of random mutagenesis techniques have been developed (see below), this problem is far from trivial. Mutations must evenly cover the region of interest, and not be biased for purely technical reasons. Favorable and unfavorable mutations may become covalently linked on the same genome, and obscure the effect of the beneficial mutation and prevent the further selection of this variant. Furthermore, the size of a library can be too small to comprehensively test all variants. It is important to realize that the library size is given not by the chemical efficiency of making a DNA library, but by providing such a library in a screenable format, which usually involves bringing it into bacterial cells. Unfortunately, most of the diversity is lost in this last step, but a solution to this problem will be discussed below.

The second challenge is to devise efficient selection techniques. A number of currently available screening technologies have been reviewed (Phizicky and Fields, 1995) and some are discussed in more detail below. In the context of using molecular evolution for understanding protein structure, folding and mobility, an indirect strategy has to be chosen: the quantity of interest has to be coupled to a selectable property of the protein. For example, ligand binding can be used as a selectable property as long as it depends on the protein being in the native state. Thus, all molecular quantities which influence the number of molecules which reach the native state (e.g. the thermodynamic stability of the protein) can be selected, by enriching those variants which outcompete the others in ligand binding. In other words, those mutants will win the race of which a higher percentage of molecules reaches or remains in the native state—as long as a direct change of the binding pocket can be excluded.

The third challenge is to make the selection strategies so simple that they can be carried out over several generations (usually called "rounds"), that is cycles of mutagenesis and selection. Multiple cycles are necessary, since it is statistically unlikely to find a collection of favorable mutations in a single round, and back-of-the-envelope calculations immediately show that it is not possible to ever make libraries large enough to contain all possible multiple mutations of a protein. Thus, a stepwise approach is needed, just as nature uses, and hence, the cycles of mutagenesis and selection must be very convenient to carry out, in order to allow true evolution over many generations.

3. LOCALIZED AND WHOLE-GENE RANDOM MUTAGENESIS

As detailed above, the evolutionary strategy consists of an alternation between random mutagenesis and phenotypic selection for the property of interest. In this section, a few random mutagenesis strategies will be briefly reviewed. We can first distinguish whether the mutagenesis is completely undirected and scattered all over the protein, or focused to particular regions in the gene (Fig. 1). The two experimental examples given in the last section will illustrate both strategies.

Which one to use obviously depends on whether the phenotypic effect is controlled by a local stretch of sequence or the global structure of the protein. Antibody combining sites (Fig. 2) are a good example for the difficulty in answering this question. Numerous studies have shown that, as one would expect from the 3D structure (Padlan, 1996), the

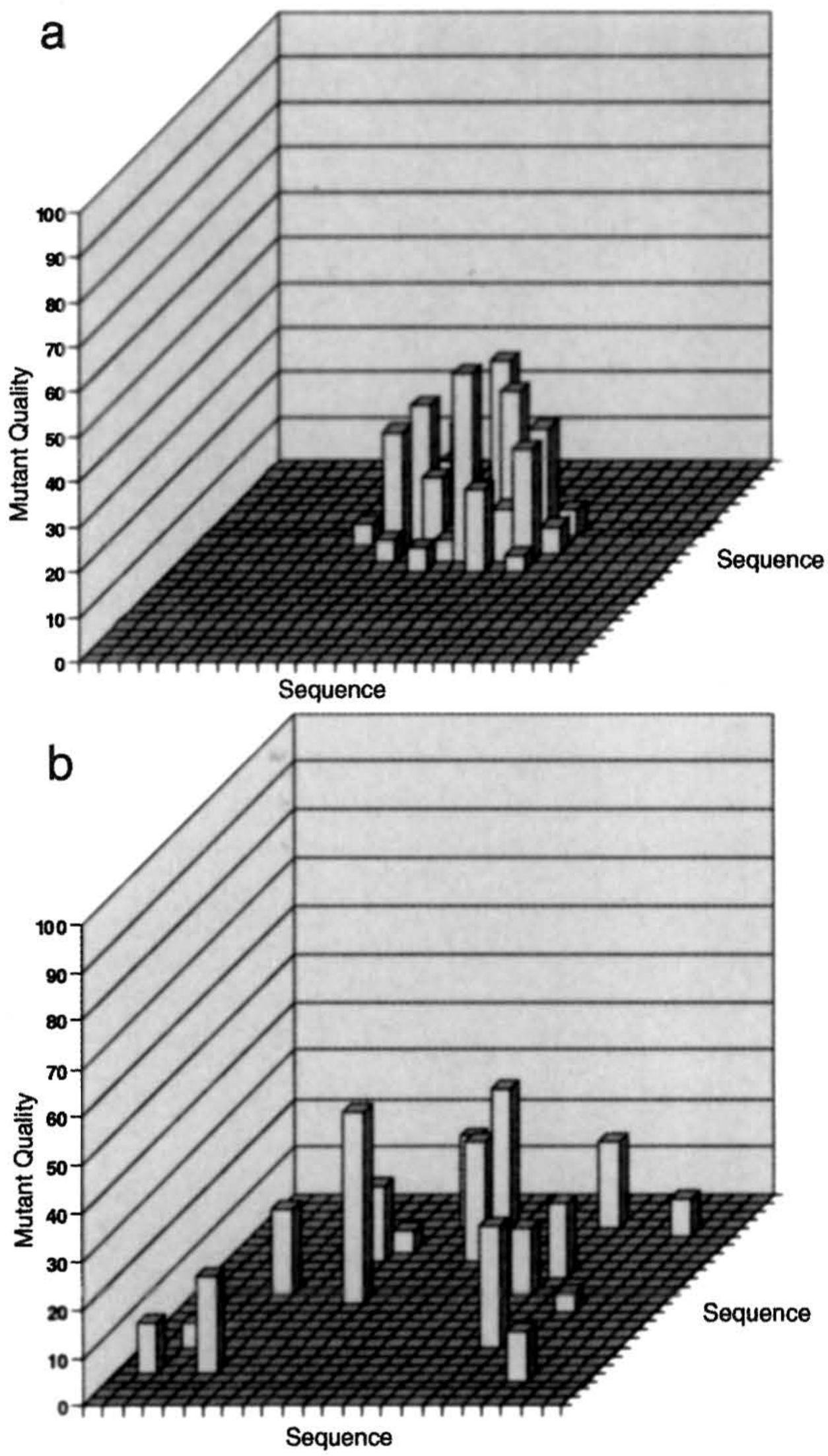


Figure 1. Localized and non-localized mutagenesis. Very schematically, two hypothetical cases are shown, how double mutations in a protein may influence its function (denoted as mutant quality in the z-axis). In (**a**), it is assumed that only a small region in the middle of the sequence is responsible for improving the function. The height of the bars symbolizes the effect of mutations at a particular position in the sequence, and only double mutations are shown (the two sequence axes). All mutations which improve function are localized in the central part of the sequence, and thus the best strategy is clearly to localize mutagenesis to this region. In (**b**), mutations all over the sequence can improve the function, and are correlated in complicated ways, that is mutations in the left part of the sequence can positively interact with those in the right part of the sequence, giving rise to bars scattered all over the 2D map. Importantly, for technical reasons, the total number of bars is about the same (representing the number of mutants which can be experimentally examined, because of restrictions in library size). Thus, in (**b**), an undirected random sampling of mutants is the better strategy than would be the comprehensive sampling of a small sequence region.

residues directly contacting the antigen, which are located in the complementary determining regions (CDRs), play the major role in providing binding energy. However, additional residues, contacting the CDRs, but not the antigen, can further improve binding by subtly changing the conformation of the CDRs. It is this "second sphere" of amino acids, whose contribution to binding is—unfortunately—beyond theoretical prediction today, as the effects are small, indirect and all interdependent on each other. Thus, antibody combining sites may be optimized in two steps: first in a "coarse" adjustment, by selecting the CDR sequences, and second in a "fine tuning", by optimizing their conformation and positioning in space, i.e., by optimizing the "second sphere". It may come as no surprise that this is exactly what happens in the affinity maturation of the antibody during the maturation of the B-cell in the immune response (Wagner and Neuberger, 1996).

In reproducing these mutations *in vitro*, one can also use localized and complete mutagenesis of the gene in question. One of the most effective methods of localized mutagenesis is to replace the CDRs by cassettes made from synthetic oligonucleotides which encode all the variability desired. Usually, phosphoramidites of mononucleotides are the synthetic building blocks for oligonucleotides. In this case, however, the mixture of codons arising from a number of mixed mononucleotides cannot be precisely controlled.

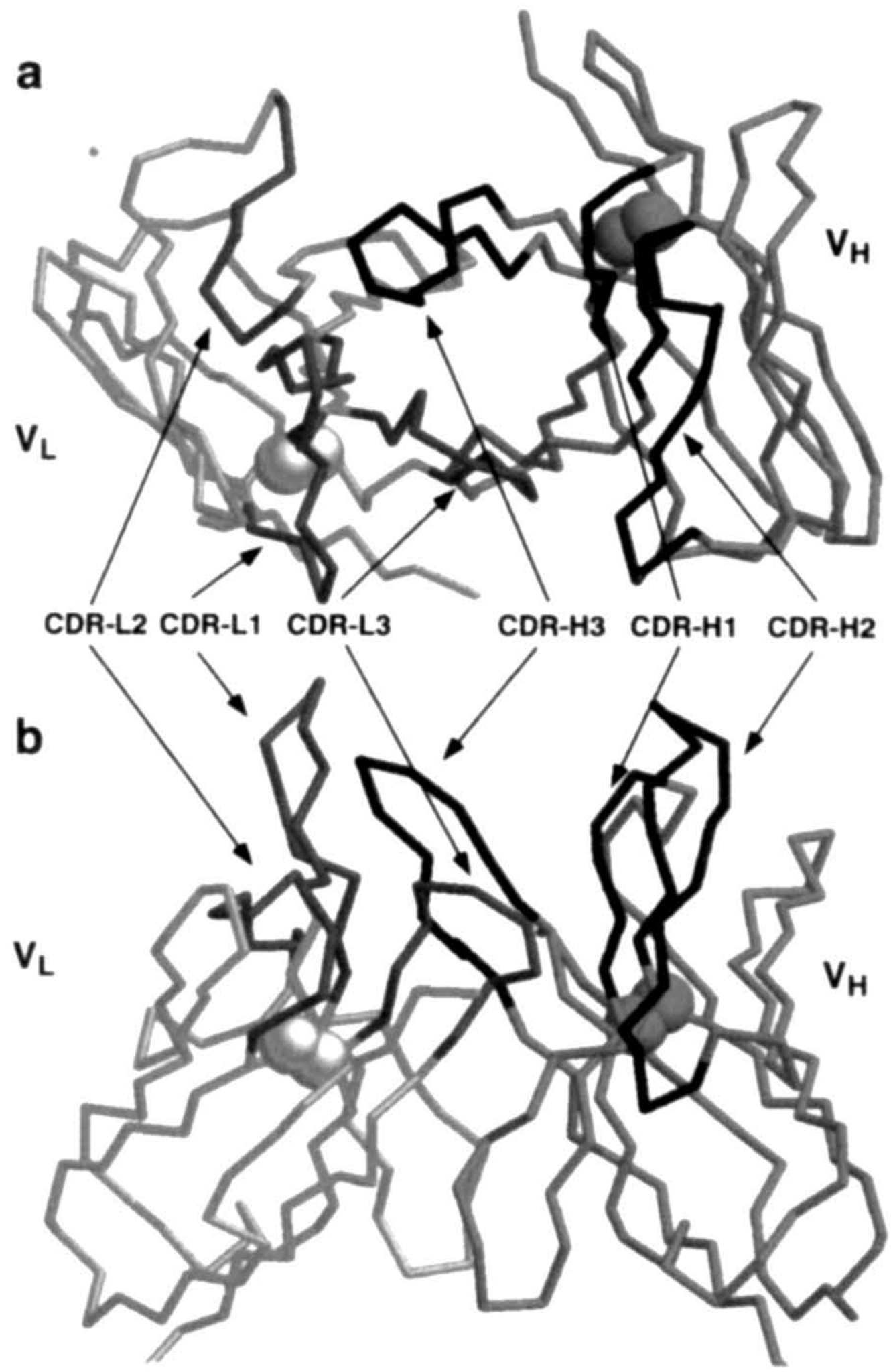


Figure 2. Structure of antibody variable domains, exemplified by the phosphorylcholine binding antibody McPC603 (PDB file 2mcp). The domains and the complementarity determining regions are labeled. The disulfide bonds in the interior of each domain are labeled. (**a**) view from the perspective of the antigen, (**b**) view from the side.

The solution consists in presynthesizing the codons for all 20 amino acids in a format compatible with a modern synthesizer (Virnekäs et al., 1994; Kayushin et al., 1996) (Fig. 3a). This means that the building blocks are phosphoramidites of whole codons and carry protecting groups just as mononucleotides would (Fig. 3a). These building blocks can then be mixed in any ratio, and only those trinucleotides can be added during the synthesis of a particular codon which are desired for the mixture of amino acids to be encoded (Fig. 3b).

To randomize whole genes, in contrast, PCR based methods are most convenient, as they are completely general and can be used in any gene. Many methods have been described for "error prone PCR" (Cadwell and Joyce, 1994), that is conditions in which the error rate is higher than normal. These methods suffer from the inherent disadvantage that either the error rate is too low and the majority of molecules will still be wild-type, or the error rate is too high, and each individual molecule is likely to contain beneficial and deleterious mutations at the same time (Fig. 3c). If several beneficial mutations are required in order to detectably improve the phenotype, it is likely that such molecules will never be found, as they always occur in the context of deleterious mutations, which are probably far more frequent.

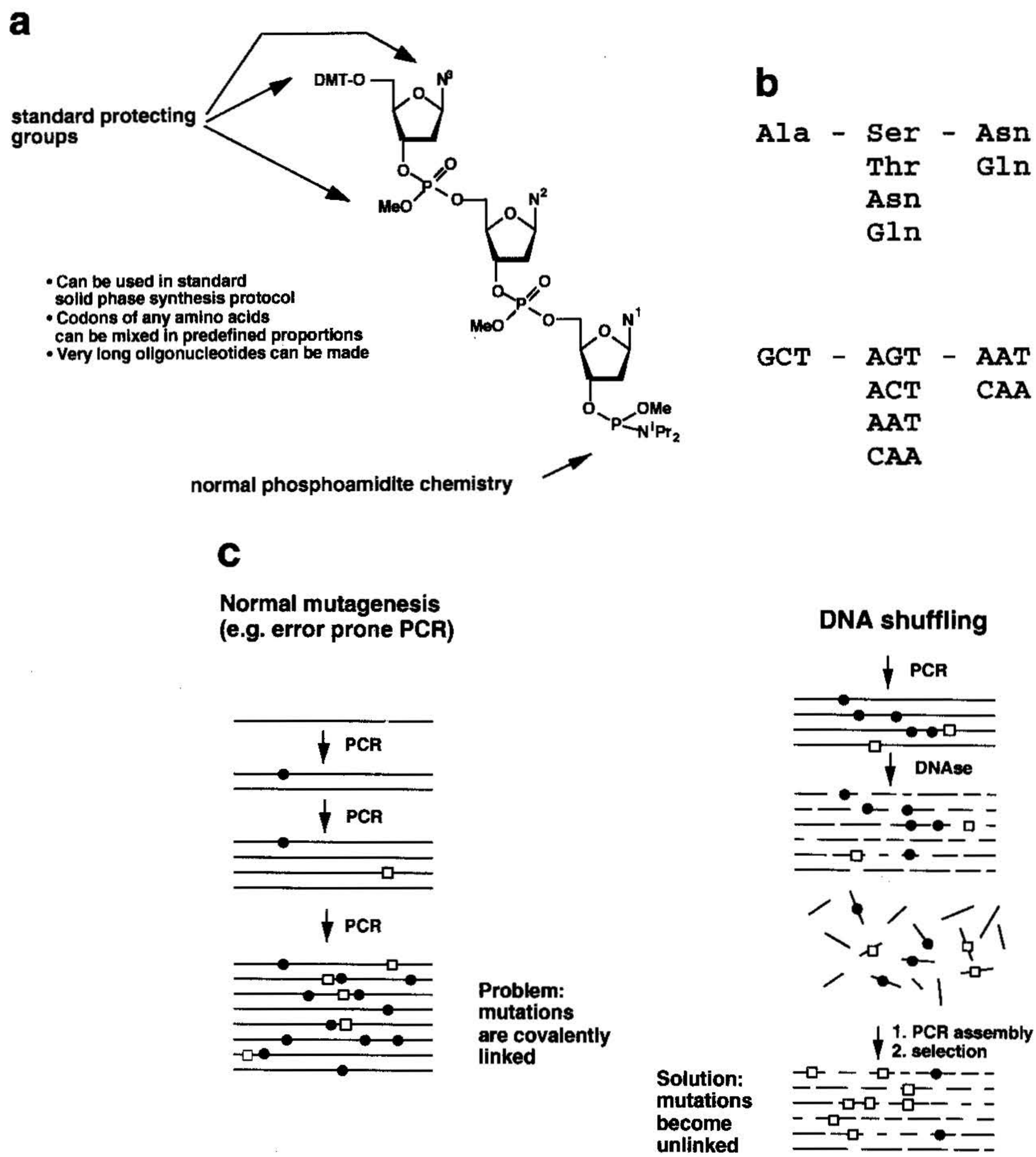


Figure 3. Mutagenesis strategies for localized and whole-gene mutagenesis. (**a**) Trinucleotide building blocks for use in oligonucleotide synthesis. Such presynthesized trinucleotides, representing codons, can be used in standard solid-phase oligonucleotide synthesis. The details can be found in Virnekäs et al., (1994). The chemical structure of the trinucleotide building block is shown, and N^1, N^2 and N^3 are bases. (**b**) Use of the trinucleotides in designing a gene with mixed codons. If a gene with the variability given in the top is desired, the codons given in the bottom need to be mixed during synthesis, using the building blocks given in (a). These mixed codon oligonucleotides are useful in the localized mutagenesis strategy schematically shown in Fig. 1a. (**c**) Mutagenesis strategies useful for undirected mutagenesis, sprinkling mutations over a whole gene of interest. On the left, error prone PCR (see text) is shown schematically. Two types of mutations are shown, favorable ones (open squares) and unfavorable ones (closed circles). In successive cycles of PCR, more mutations of each are introduced, and usually molecules will contain some of either type. Thus, the beneficial effect of the favorable mutations can be completely obscured by the presence of unfavorable ones. On the right, DNA shuffling according to Stemmer (1994a) is shown. A DNase step breaks up the DNA to small pieces, and PCR is used to reassemble the gene. Thereby, mutations are crossed, and genes with largely favorable mutations will be obtained which can be enriched by selection.

Stemmer (1994a) described an elegant solution to this dilemma. By carrying out PCR with an enzyme without proofreading functions, such as the *Taq* polymerase from *Thermus aquaticus*, a certain number of errors are introduced. By then digesting the gene with DNAse into small pieces and PCR assembling the gene from the pieces, the mutations are "crossed", just as if they were located on individual chromosomes (Fig. 3c), even

though they belong to the same open reading frame of the protein. Thereby, beneficial mutations can be isolated from deleterious ones by selection (next section), and evolution is much faster than with other methods (Stemmer, 1994a,b).

4. SELECTION TECHNOLOGIES

4.1. General

The common requirement of all selection techniques is to provide such a linkage and physically connect the phenotype with the genotype of the protein (Phizicky and Fields, 1995). That is, if a particular protein molecule has more desirable properties than the rest, because it carries one or more useful mutations, there must be a way of amplifying and immortalizing this particular sequence. Hence, one must achieve a 1:1 physical linkage of the DNA sequence (the genotype, which can be inherited) and the protein sequence (the phenotype, which provides the properties). If there was a pool of protein molecules and a pool of DNA sequences, but no such direct physical linkage, there would be no way that the particular DNA sequence belonging to the improved protein could be identified and amplified.

A whole array of techniques has been developed to select for molecular interactions, such as phage display (Smith, 1985; Dunn, 1996), the yeast two-hybrid system (Bai and Elledge, 1996; Warbrick, 1997), the peptides-on-plasmids system (Cull et al., 1992), ribosome display (Mattheakis et al., 1994; Hanes and Plückthun, 1997), yeast surface display (Boder and Wittrup, 1997) and bacterial display (Georgiou et al., 1997). To select for structural features, however, the selection has to be quite resistant to enriching "sticky" molecules and must be convenient enough to be applied over multiple rounds, in order to enhance very small selective advantages. In this article, three methods will be discussed in more detail which have been shown to be particularly useful in the field of molecular evolution, namely phage display, selectively infective phages and ribosome display.

4.2. Phage Display

The most popular technique used in this field today is phage display (Smith, 1985; Scott and Smith, 1990; Winter et al., 1994; Dunn, 1996; Cortese et al., 1996). It relies on fusing the protein of interest to the minor coat protein of the phage, the gene3 protein (g3p). This protein consists of three domains of 68 (N1), 131 (N2) and 150 (CT) amino acids, connected by glycine-rich linkers of 18 (G1) and 39 (G2) amino acids, respectively (Fig. 4). The first N-terminal domain, N1, is thought to be involved in penetration of the bacterial membrane, while the second N-terminal domain, N2, may be responsible for binding of the bacterial F-pilus (reviewed in Spada et al., 1997). Recently, the structure of the N1-N2 domain complex has been solved at very high resolution (Lubkowski et al., 1998).

In phage display, a ligand (e.g. an antigen) is immobilized and a collection of binding proteins (e.g. antibodies) are displayed on the phage, that is, provided as fusions to the minor coat protein g3p (Winter et al., 1994). The essential trick is now that the genetic information of the displayed protein is contained within the phage DNA in the interior of the same phage particle (which is packed from the bacterial cell which makes the phage and produces the proteins) and thus, physically connected to the expressed protein. When the collection of antibody mutants is then passed over the antigen, only those which recognize

the antigen will bind. They can then be eluted from the antigen and used to infect cells, and thereby the useful information, namely which sequence bound the antigen, can be amplified and deciphered.

This method is easy to understand when the protein variants differ in the binding site for the antigen. However, it also works when the collection of protein species differs in more global parameters, such as stability (Jung and Plückthun, 1997). The reason is that the experiment is carried out with populations of molecules: If a protein is unstable, only

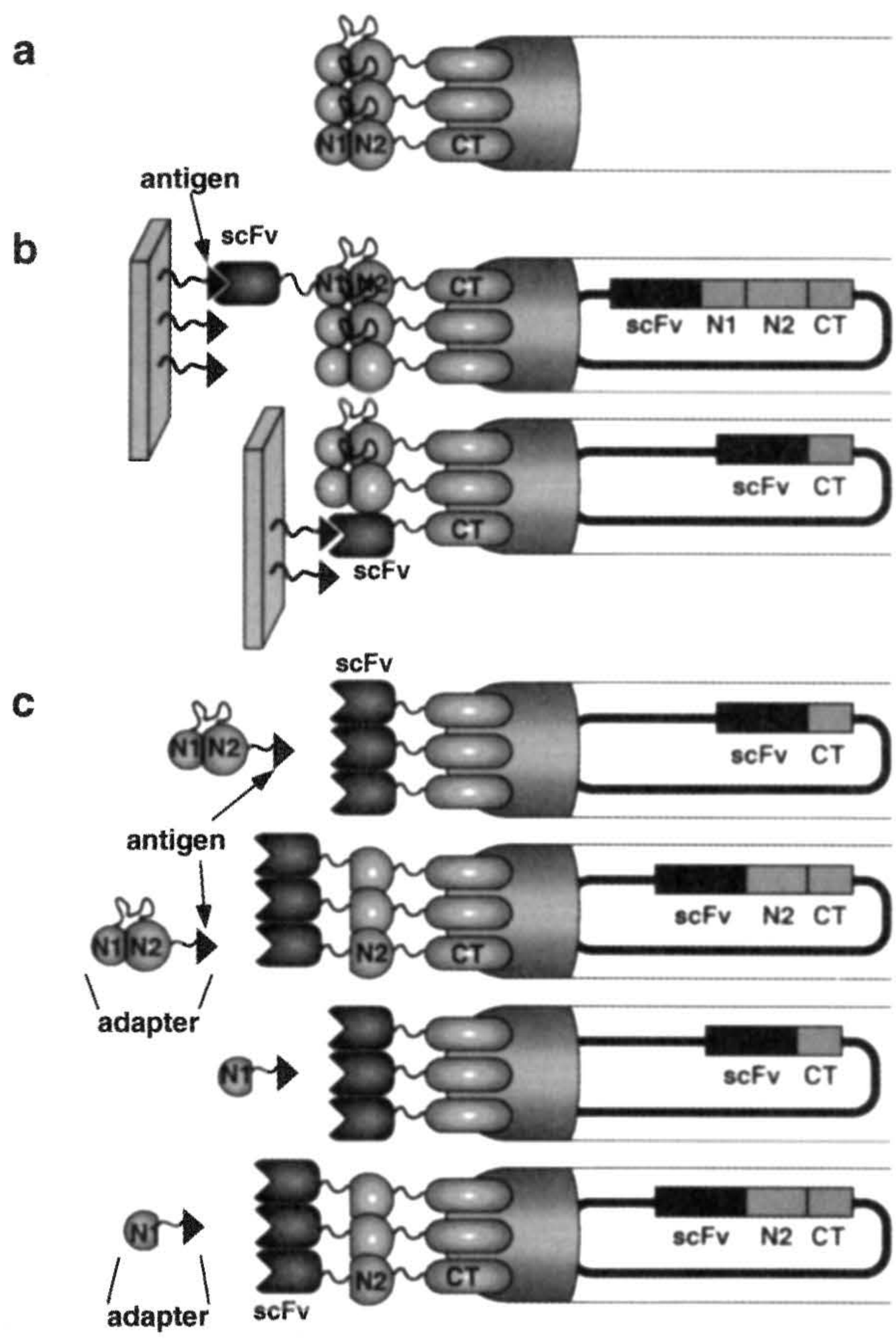


Figure 4. (**a**) On the top the w.t. phage is shown which contains g3p at one tip of the phage. The domain structure of g3p is indicated (for details, see text). (**b**) In conventional phage display, fusions of the protein of interest to the N-terminal domain (long fusions) or to the C-terminal domain (short fusion) are expressed together with a helper phage, which provides the g3p w.t., needed for infection. The desired phages are obtained from the mixture by "panning", that is binding to a solid surface, as schematically indicated. Each phage contains the DNA encoding the fusion protein. (**c**) SIP phages do not contain the N1 domain, which is absolutely needed for infection. N1 is provided by the adapter protein and must be linked to the phage by the cognate interaction. There are four different possibilities for constructing SIP phages (top to bottom), which differ in the phage either containing N2 or not, and in the adapter either containing N2 or not. The detailed consequences of these constructs are discussed elsewhere (Krebber et al., 1997).

few molecules from the population will be in the native state. If a phage displaying such a protein has to compete with one where a much larger percentage has reached the native state, the latter will bind much more frequently and outcompete the former. Thus, the selection is indirect and can use the "percentage of native molecules" as the ultimate parameter which is selected for.

The trouble with phage display is that it does not always work as smoothly, as the above description would suggest. The reason is that there are many ways how a phage can "stick" to the immobilized ligand: unfortunately, an unfolded protein or a partial sequence usually exposes a number of hydrophobic residues which stick strongly to the surface in question, but not because the molecules specifically recognizes the ligand (the antigen). Thus, several "rounds" are necessary, i.e., cycles of enrichment, in which the "right" binders are enriched over the "wrong" ones. However, the method is still not always successful. Therefore, two alternative methods, SIP and ribosome display, have been developed, described in the next two sections.

4.3. Selectively Infective Phage

The selectively infective phage (SIP) (Krebber et al., 1993; 1995; 1997; Duenas and Borrebaeck; 1994; Gramatikoff et al., 1994; Spada and Plückthun, 1997; Spada et al., 1997) is related to phage display technology using filamentous phages as described in the previous section. In contrast to phage display, the phage particles are rendered non-infective in SIP by disconnecting the N-terminal domains (N1 or N1-N2) of the phage g3p coat protein (Fig. 4b)–those involved in docking and bacterial cell penetration–from the C-terminal domain (CT) which caps the end of the phage. The N- and C-terminal domains are then each fused to one of the interacting partners being studied. One partner is thus displayed on the phage surface associated with CT, while the other is genetically fused or chemically coupled to the N-terminal domain(s), thus constituting a separate "adapter" molecule. Only when the specific protein-ligand interaction occurs between the partners is the g3p reconstituted and does the phage particle regain its infectivity, and the genetic information of a successful binder is propagated. In the absence of a cognate interaction, no infectivity is observed, demonstrating that the N-terminal domains really must be physically coupled to the phage for infection. In contrast, in phage display the whole phage binds and can be eluted from an immobilized target molecule, whereupon it can inject its DNA into bacterial cells and be amplified, since it remains infective.

SIP experiments can be carried out *in vivo* or *in vitro* (Fig. 5): in the former approach the phage or phagemid encodes both interacting molecules (e.g. antibody-antigen) in the same vector (Fig. 5a), and the protein-ligand interaction occurs in the bacterial periplasm; the infective phage particle is extruded from the cell and can be harvested and used to infect different bacteria. The simplicity of design of the experiment is counteracted to some extent by the lack of experimental control over parameters such as concentrations, incubation times and the infection process itself.

These factors can be controlled in the *in vitro* SIP experiment (Fig. 5b). In this case, the adapter protein is prepared from a separate *E. coli* culture, either by secretion or by *in vitro* refolding. This approach broadly extends the range of applications of SIP, as, in addition to genetically fused peptidic antigens, non-peptidic antigens can now be chemically coupled to the N-terminal domains. The adapter is then incubated with the non-infective phage library, and added to the bacteria.

A requirement of SIP appears to be a high affinity between protein and ligand, and all of the systems studied thus far have had dissociation constants of at least 10^{-8} M. It is

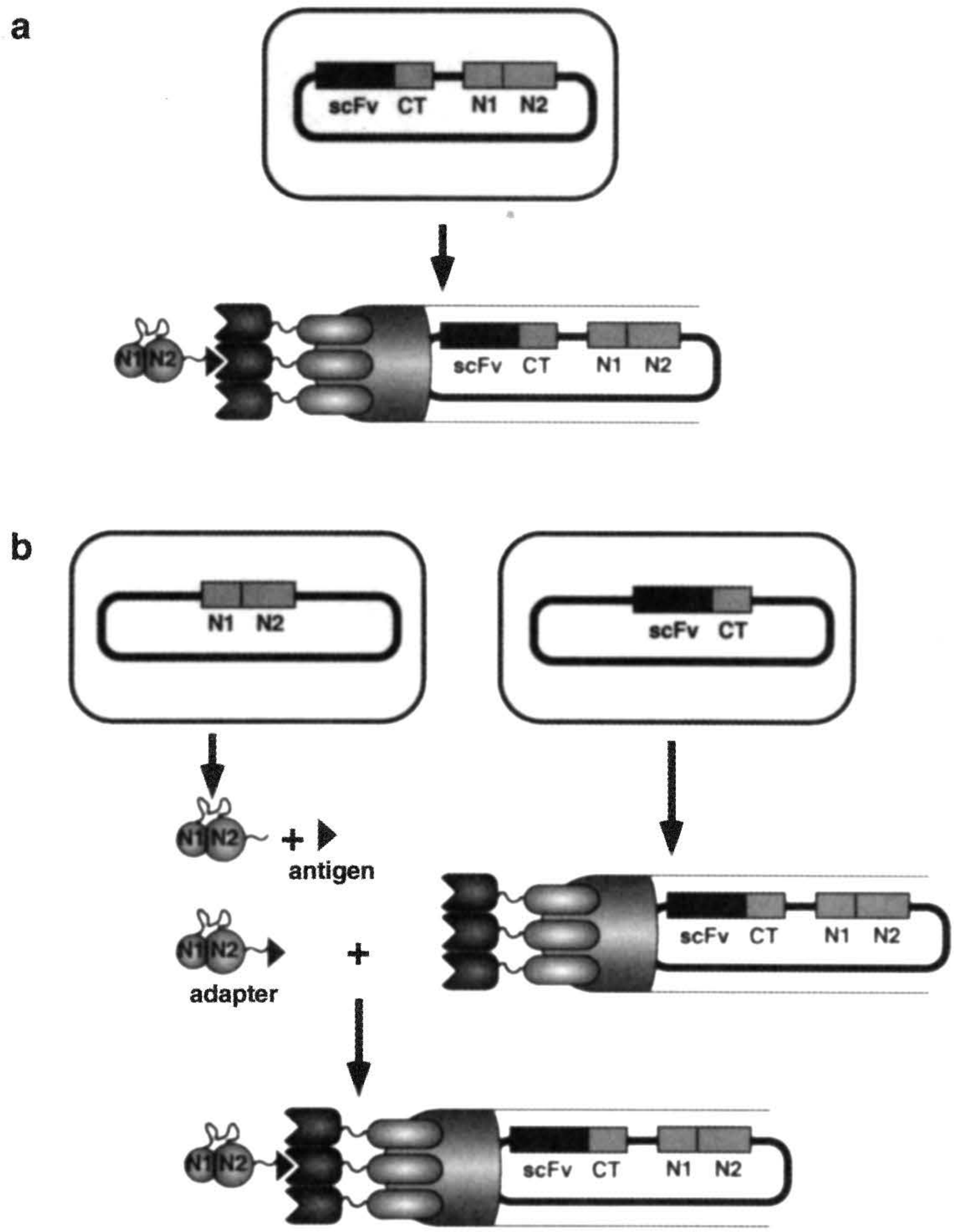


Figure 5. *In vitro* SIP vs. *in vivo* SIP. (**a**) The phage can encode both the ligand (fused to N1-N2 in this example) and the receptor (shown as a scFv fragment of an antibody in this example, fused to CT) (*in vivo* SIP). In this case, the *E. coli* cell produces an infective phage, if the ligand and receptor recognize each other, as the interaction already occurs in the bacterial periplasm upon budding of the phage. (**b**) Alternatively, a separate cell can be used to produce the adapter protein, which can be chemically coupled to the ligand (shown on the left), or provided as a fusion protein (not shown, analogous to (**a**)). The phage (as shown on the right) then does not encode the adapter and needs to be mixed *in vitro* with the adapter protein (*in vitro* SIP).

not yet clear, however, what the minimum affinity necessary for significant infectivity will be. Further studies will have to further quantitatively investigate the dependence of infectivity on properties of the proteins such as thermodynamic stability and expression rate.

No solid phase interaction with any support is necessary in SIP, limiting the potential of the occurrence of non-specific interactions and eliminating the need for elution. In addition, the low background infectivity observed with some of the SIP systems demonstrates that SIP can be an extremely effective and highly specific method of selecting for cognate interaction events. Furthermore, the exciting possibility of simultaneous screening

of two interacting libraries by *in vivo* SIP might broaden the ability to identify interacting receptor-ligand pairs.

4.4. Ribosome Display

The development of the technology of "ribosome display" (Fig. 6) has been an effort to overcome a shortcoming all other technologies have in common, namely a limited library size (Fig. 7). The method was first developed for short peptides (Mattheakis et al., 1994), and then for whole proteins (Hanes and Plückthun, 1997). The trick is to carry out the selection entirely *in vitro*, without the use of any cells or phages at all. The concept is actually very simple. A pool of protein molecules is translated on ribosomes *in vitro* such that three conditions are fulfilled (Hanes and Plückthun, 1997). First, the protein must fold to is correct three-dimensional structure directly on the ribosome. Second, the protein must not leave the ribosome after the synthesis is finished. Third, the RNA (which carries the genetic information for the protein) must not leave the ribosome, either, after the synthesis is finished. Thereby, the phenotype and the genotype become linked *in vitro*—on the ribosome that makes the protein in the first place.

Before discussing the methods used to effect folding of the protein on the ribosome and preventing the dissociation of mRNA and protein, the advantages of this system shall be summarized. Since all steps occur *in vitro*, no transformation is necessary (Fig. 7). In all other selection systems, the DNA library needs to be taken up by bacteria. Usually this is done by electroporation (Dower and Cwirla, 1992), and while this method has been continuously optimized, only a tiny fraction of a typical ligation reaction is taken up by the cells. Typically, ligation reactions will be carried out with 10^{11}–10^{12} molecules of vector DNA, but in a single electroporation cuvette, typically only 10^{7}–10^{8} molecules will be taken up to transform the bacteria—a tiny sample of the variety originally present. In the ribosome display technology, there is no transformation, and all the original library size is directly used in screening.

The ribosome display technology (Fig. 6) works by a modification of standard *in vitro* translation. A library of genes for protein molecules is amplified by PCR, and thereby a promoter is added at the 5′ end. To provide resistance against nucleases, the mRNA also contains hairpins at both the 5′ and the 3′ end. Importantly, there is no stop codon present. The DNA library is transcribed, and the mRNA is isolated. This separate transcription and translation is advantageous as it allows different conditions for transcription and translation. Particularly, RNA polymerases require DTT, and the use of this compound during translation would largely prevent disulfide formation in the nascent protein. However, disulfides seem to be needed e.g. for the functional expression of antibody fragments (see also below). Therefore, translation is carried out separately from transcription, and the redox potential for the translation can be adjusted as needed and carried out in the presence of protein disulfide isomerase (PDI), which is essential for disulfide formation (Ryabova et al., 1997; Freedman et al., 1994). Furthermore, a whole series of combinations of molecular chaperones has been investigated and shown to further increase the yield of folded protein (Ryabova et al., 1997). The translation is then stopped by cooling on ice, and in the presence of very high Mg^{2+} concentrations, which stabilize the ribosomes against dissociation and thereby against release of mRNA and protein.

The method is made more efficient by also paying attention to the translation conditions. Vanadate transition state analog inhibitors of RNase are used. Additionally, an inhibitor of the *E. coli* 10Sa RNA (the product of the *ssrA* gene) is added. This RNA is specifically designed by the cell to remove proteins and mRNA which remain associated

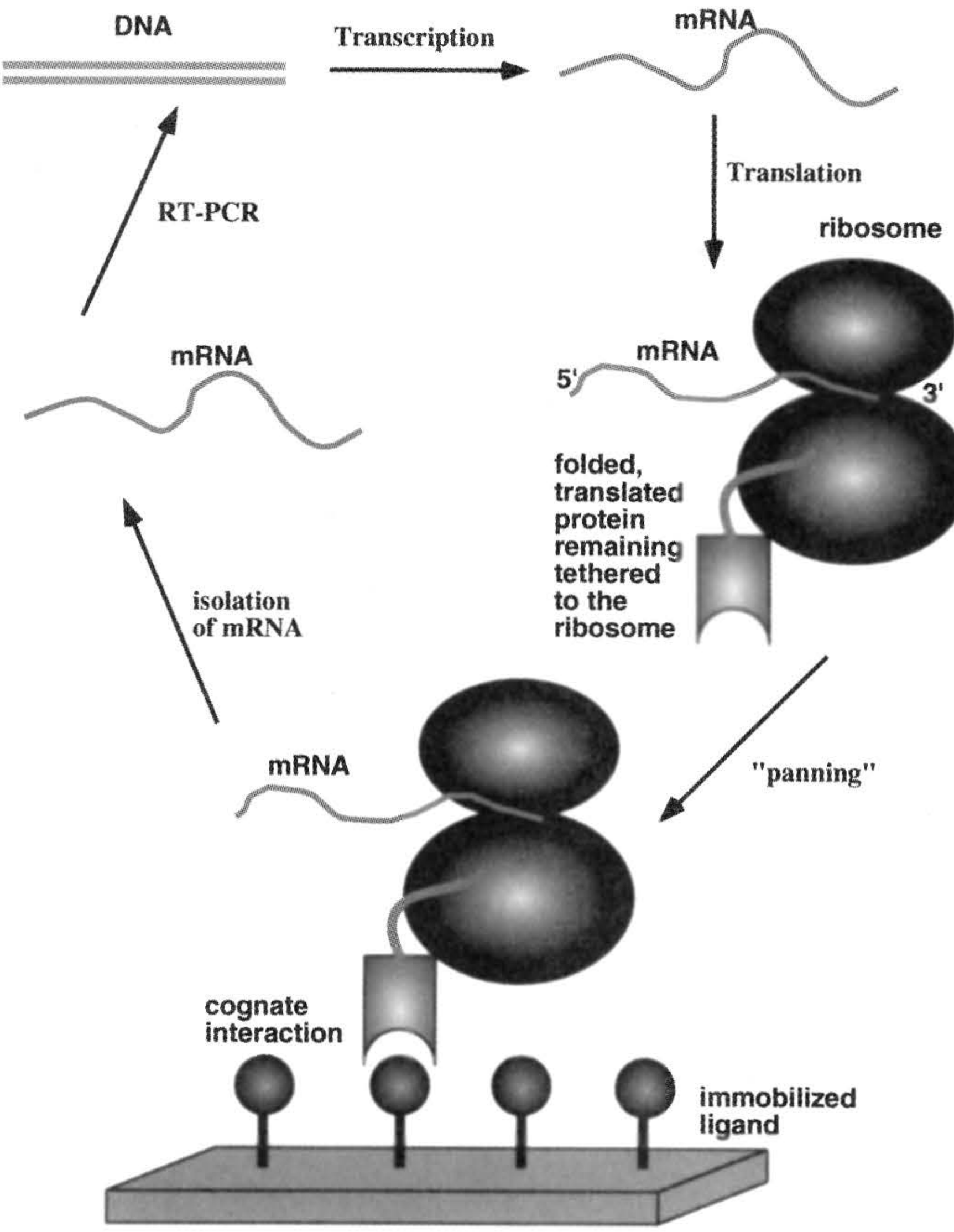


Figure 6. Principle of *in vitro* ribosome display for screening native protein (scFv) libraries for ligand (antigen binding. A DNA scFv library is first amplified by PCR, whereby a T7 promoter, ribosome binding site and stem oops are introduced, and then transcribed to RNA (top). After purification, mRNA is translated *in vitro* in an *E coli* S-30 system in the presence of different factors enhancing the stability of ribosomal complexes and improvin the folding of the scFv antibodies on the ribosomes (for details, see Hanes and Plückthun, 1997). Translation then stopped by cooling on ice, and the ribosome complexes are stabilized by increasing the magnesium concen ration. The desired ribosome complexes are affinity selected from the translation mixture by binding of the nativ scFv to the immobilized antigen (bottom). Unspecific ribosome complexes are removed by intensive washing. Th bound ribosome complexes can then either be dissociated by EDTA or whole complexes can be specifically elute with antigen. RNA is isolated from the complexes (left). Isolated mRNA is reverse transcribed to cDNA, an cDNA is then amplified by PCR. This DNA is then used for the next cycle of enrichment, and a portion can b analyzed by cloning and sequencing and/or by ELISA or RIA.

with the ribosome, when the mRNA does not have a stop codon (Keiler et al., 1996) However, since this association is the very basis of ribosome display, the 10Sa RNA mus obviously be eliminated.

In studying the method first with a simple example of two antibodies, it was foun that it is extremely efficient. The mRNAs were mixed at a ratio of $1:10^8$ and the dilute on was enriched over 5 cycles (Hanes and Plückthun, 1997). More exciting was the discover that, by the use of *Taq* polymerase, which has an appreciable error rate of about 1 in 10 nucleotides (Keohavong and Thilly, 1989), mutations in the selected molecules wer

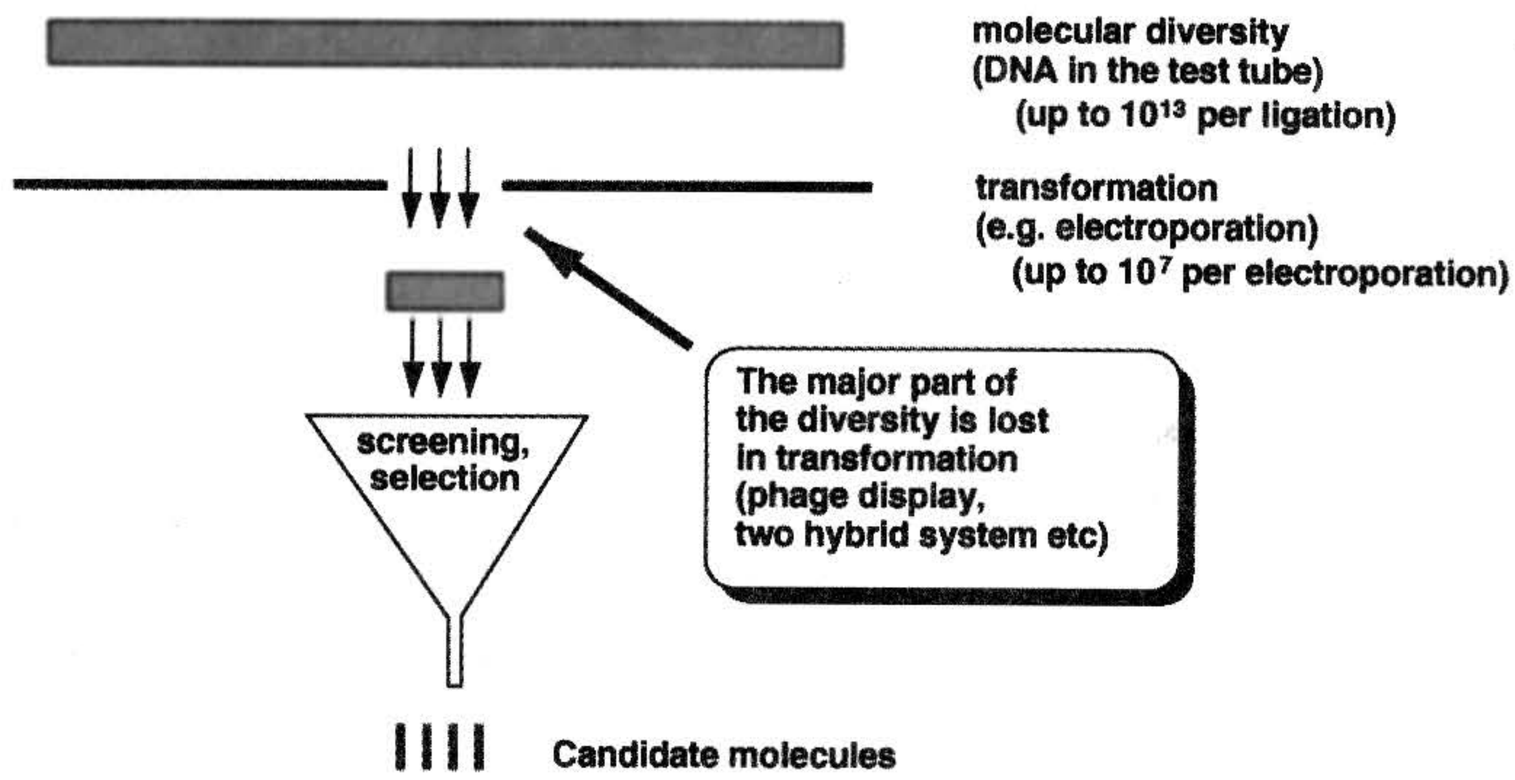


Figure 7. Schematic representation of the transformation problem. In the creation of a library *in vitro*, a large number of diverse molecules is created and—usually—eventually ligated to a plasmid. Upon transformation of cells (usually, electroporation of *E. coli*), only a small fraction of this library is actually taken up by the cells and subjected to phage production and panning. Thus, only a small fraction of the library actually takes part in the selection process. This problem can be alleviated by repeating the electroporation process many times, until the desired library size is achieved, or by using cell-free panning strategies.

found: the population has clearly evolved to remain compatible with binding. It is this easy introduction of mutations which is a most exciting prospect, as it may allow us to increase the stringency of selection and thereby truly improve a phenotype in real time. Since all steps occur *in vitro*, mutagenesis and selection is much faster achieved than in the phage based methods detailed above. Nevertheless, some proteins are more easily enriched than others, and the molecular reasons for this are still being investigated.

5. CASE STUDY I: A STABLE DISULFIDE-FREE ANTIBODY SCFV FRAGMENT BY DIRECTED EVOLUTION

Evolutionary technology has been applied to a number of different problems in protein optimization (summarized in the Introduction). Here, the power of this approach for studying problems of protein structure and function shall be exemplified with two studies on antibody structure.

In the first study, an evolutionary approach was used to generate a disulfide free antibody single-chain Fv (scFv) fragment (Proba et al., 1998). In a scFv fragment, the variable domain of the heavy chain (V_H) is connected by a flexible linker to the variable domain of the light chain (V_L) (reviewed by Huston et al., 1993). The Fv or scFv fragment is thus the smallest unit still containing the full binding site of the antibody.

The antibody domain consists of a beta sandwich, with five and four strands making up the two sheets (Fig. 2). There is a very conserved disulfide bond, connecting strands b and f. About 99.5% of all sequences (murine and human) in the Kabat database and all the germline genes of the human antibodies (Vbase, version 1997, http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html) contain both of the cysteines making up this disulfide bond.

In the cytoplasm of the cell, the redox potential is reducing, and no disulfide bonds can form (Gilbert, 1990). Thus, when an antibody is expressed in the cytoplasm of a cell, its disulfide bonds do not form, and most frequently the native state is not stable. There are many pieces of evidence that the disulfide bond is critical for stability. First, refolding

experiments in the presence of reducing agents and directed mutagenesis experiments have been used to replace the cysteines against other hydrophobic residues in many combinations. The stability is severely affected and sometimes, a stable protein is not obtained at all (Goto and Hamaguchi, 1979; Glockshuber et al., 1992). Nevertheless, a few natural antibodies exist (Rudikoff and Pumphrey, 1986) which are missing one of the cysteines, which must have been lost by somatic mutations occurring in the maturation of the B-cell. Since at least a few of these antibodies have been found to be functional, in these cases the stable structure seems to be reached anyway, even in the absence of the stabilizing disulfide bond. It appears, therefore, that additional stabilizing mutations can compensate for the loss of the contribution of the disulfide bond to stability. Indeed, in a study of a V_L domain, stabilizing mutations were shown to be able to partially compensate for the loss of the disulfide and lead to a protein of reduced stability (Frisch et al., 1994, 1996), but normal structure (Uson et al., 1997).

For a number of applications, it would be of great interest to construct antibodies which do not need the disulfide bond for stability, and can be expressed inside the cell as "intrabodies" (Richardson and Marasco, 1995; Biocca and Cattaneo, 1995). For example, in the various genome projects, an enormous number of new proteins will be described for which no function will be known, and it would be very helpful to establish a function by titrating out the protein in question and observing the effect is has on cellular functions. Since a recombinant antibody against the recombinant protein can be obtained very quickly (Winter et al., 1994), it would be very attractive to use such an antibody for this purpose. Thus, in many cases, the antibody would have to be functional in the cytoplasm. More ambitiously, the antibody might be a catalytic one (Lerner et al., 1991), and a functional cytoplasmic expression might be useful for integrating the new reaction into general metabolism to either select for higher activities or even, in the very long run, redirect metabolism into new directions to produce new metabolites and new pharmacophores in plants or fungi. Furthermore, cytoplasmic and nuclear expression may also be used to couple antibody antigen interaction to transcription, to either obtain a readout system by generating a color or light as a consequence of the cognate interaction or to turn on a selectable gene as a consequence. Thus, there is clearly sufficient motivation to generate functional disulfide-free scFv fragments.

Directed evolution was used to achieve this goal (Proba et al., 1998). The strategy was to select for stabilizing mutations anywhere in the structure which might overcome the loss of the disulfide bond (Fig. 8). As a starting protein, an antibody was used in which the disulfide bond in the heavy chain was missing already. In this case, Cys H92 is replaced by Tyr in the natural protein, and this protein had been shown before to be functional (Rudikoff and Pumphrey, 1986). In order to "gently" apply the evolutionary pressure, the two domains V_H and V_L were first independently optimized, then combined under further mutagenesis. Thus, the best combination of cysteine-free domains was selected and further mutations were introduced to compensate for the loss of the disulfide.

In a first set of experiments, the missing cysteine was restored in the heavy chain, such that this domain carried a disulfide bond, and in the light chain, the positions of the cysteines were randomized to Ala, Val, Leu, Ile and Phe (Fig. 8). In a parallel set of experiments, the position H22 in the V_H domain was randomized to Ala, Val, Leu, and Ser, position H98 already being tyrosine. Both libraries were screened by classical phage display, and only the combinations in Fig. 8 (third line) survived the selection.

The two sets of disulfide free domains were then combined to identify proteins which were functional in the absence of any cysteines in any domain. However, the selected domains were not simply ligated, but a DNA shuffling (Stemmer, 1994a) was carried out. This

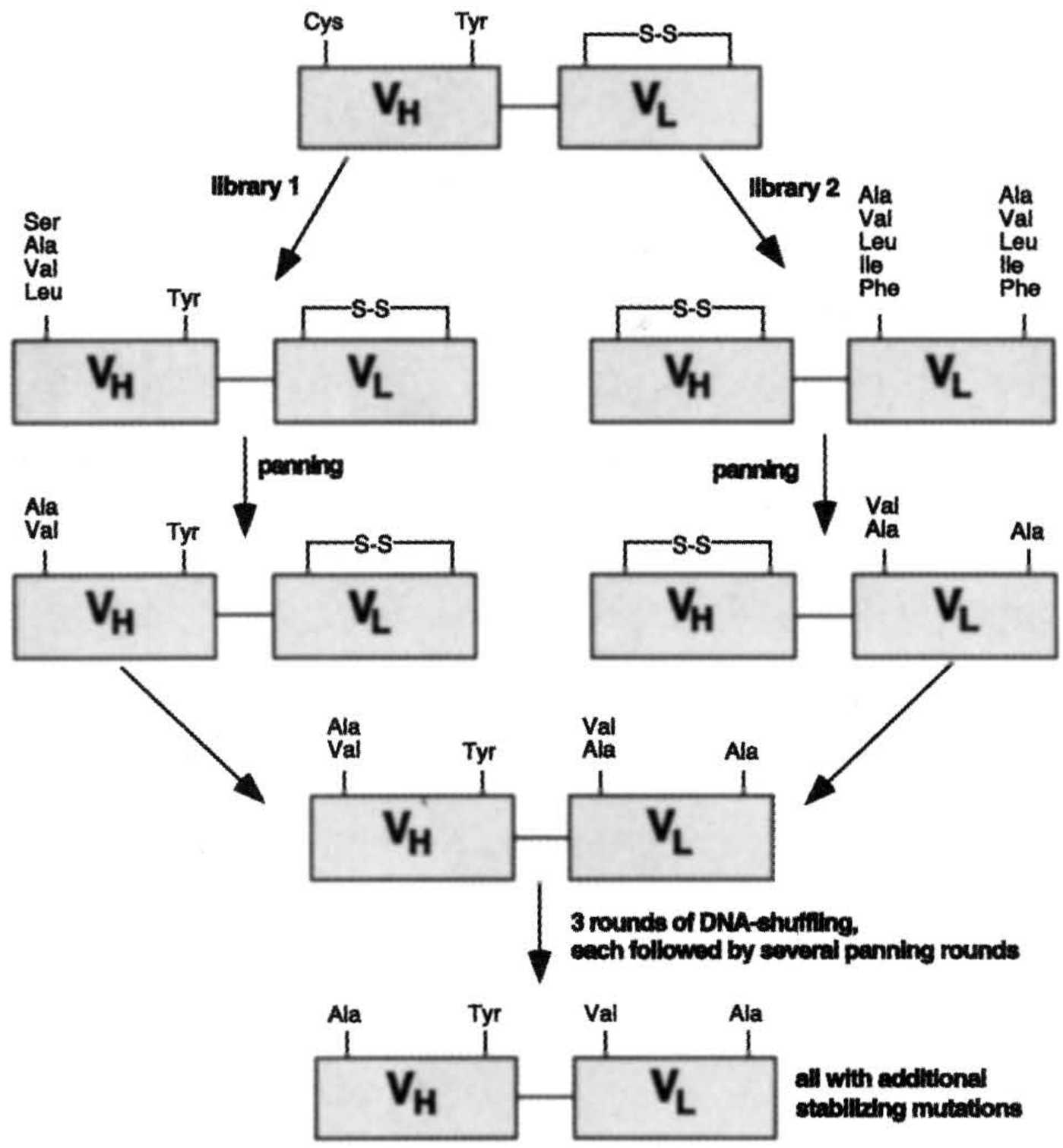


Figure 8. Schematic description of the directed evolution of disulfide-free scFv fragments. As a model system, the scFv fragment ABPC48 (naturally missing the disulfide in V_H, top line) was used. Two libraries were constructed, each restoring the disulfide in one domain and randomizing the cysteine positions in the other. The selected variants were recombined and subjected to DNA shuffling (Fig. 3c), and the final version, not containing any cysteines, was only found in the presence of additional mutations, but is practically unable to fold in the absence of further stabilizing mutations. (For details, see Proba et al., 1998).

was done to introduce additional mutations which may stabilize the molecules. Over several rounds of shuffling and selection, a range of new variants were obtained (Proba et al., 1998). The conditions for screening were slowly increased in stringency.

The result of the directed evolution experiment was a series of molecules, able to bind the antigen and not containing any cysteines. Importantly, the control molecule which only contained the cysteine replacements and not any of the additional mutations could not even be produced to test its stability (Proba et al., 1998). A number of different mutations were thus selected which all improve stability, but only by small increments. However, the combination of these mutations in the same molecule, which is one of the hallmarks of the evolutionary approach, solved the problem of making a stable, disulfide-free scFv fragment.

Some of the mutations can now be rationalized. For instance, the mutation LysH66Arg (Fig. 9) can be modeled, since this residue occurs naturally in human antibodies. It forms an ideal interaction with Asp-H86, with the two guanidino NH-groups of Arg forming hydrogen bonds to the two carboxylate oxygens of Asp. This salt bridge is quite buried, increasing its stabilizing effect compared to that of a solvent-exposed ion pair. The elucidation of the exact mechanism of how other mutations improve the activity of the protein clearly requires further studies, such as e.g. for the stabilizing mutation AsnH52Ser (Fig. 9).

Other mutations may have been selected for the efficiency of folding, and not thermodynamic stability. The ability of antibody mutants to avoid aggregation appears to be not mechanistically linked to thermodynamic stability (Knappik and Plückthun, 1995; Nieba et al., 1997; Jung and Plückthun, 1997), and it is possible that some mutants of this type have been selected.

In conclusion, it can be seen that the evolutionary approach selected a series of very subtle improvements which would have been extraordinarily hard to predict, but are now ideal objects for further studies. Thus, while the same amount of physical chemistry needs to be carried out to increase the understanding of the effect of the mutations as would have been necessary in a site directed mutagenesis approach, it is done with a much more relevant set of mutants, namely those which all lie on a path to higher stability and a higher yield of folding.

6. CASE STUDY II: THE KINK IN THE FIRST STRAND OF ANTIBODY LIGHT CHAINS

Another example which was studied with directed molecular evolution, this time using SIP (see above), is the optimal configuration of the first strand of the β-sandwich. A

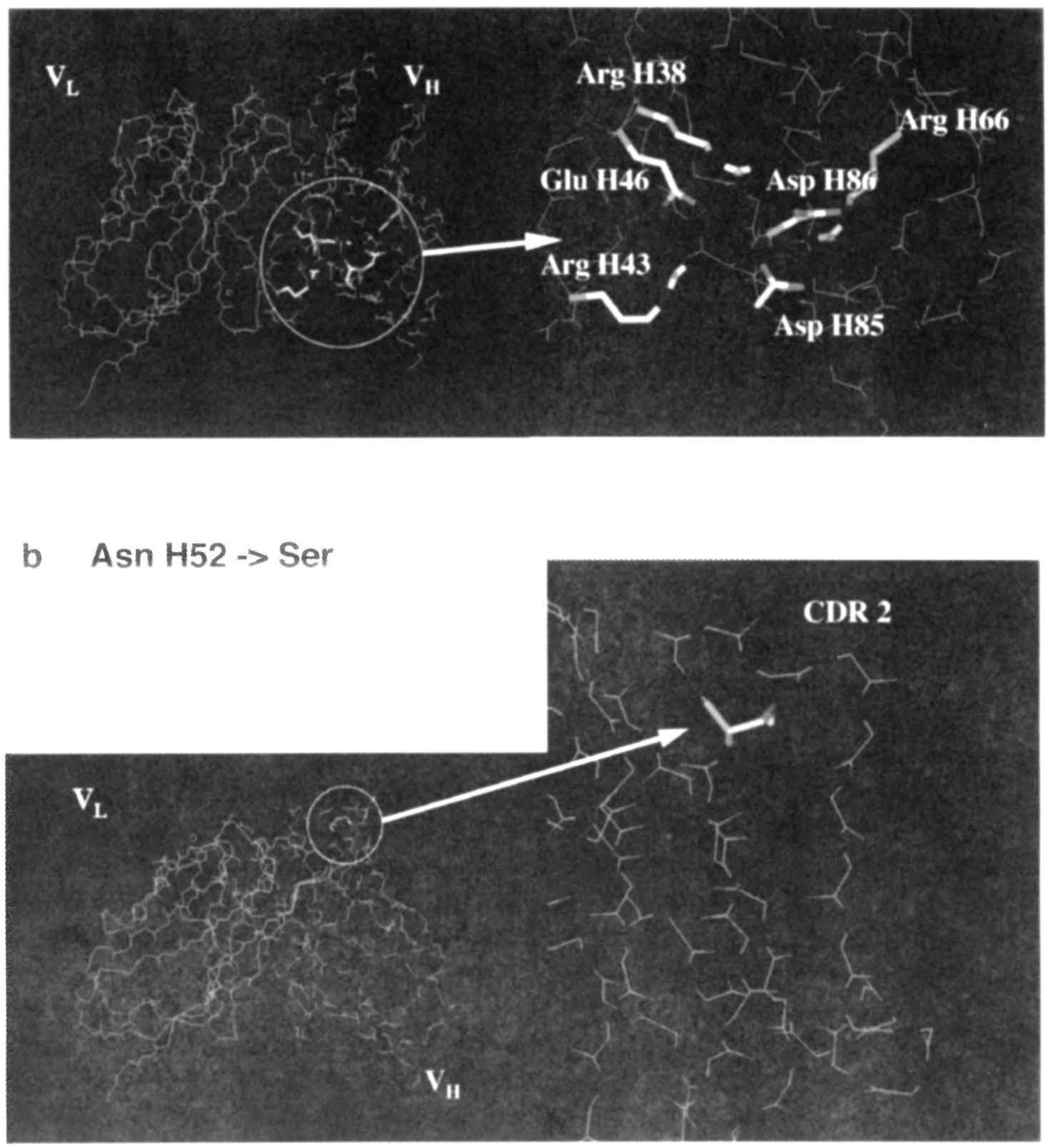


Figure 9. Two of the selected mutations are shown, which have been demonstrated to stabilize the protein, and are essential for the disulfide-free protein to remain stable. (For details, see Proba et al., 1998).

hallmark of the immunoglobulin variable domain is the first β-strand, which is interrupted. The first part of the strand makes hydrogen bonds, typical of an antiparallel arrangement, to the outer β-sheet of the V_L/V_H heterodimer. Next follows a kink, whose presence but not sequence is conserved in antibody variable domains, and then another strand which makes hydrogen bonds to the inner β-sheet of the domain. Thus, this kink is essential in closing off the sides of the sandwich to solvent, and it thus helps to connect the two sheets (Fig. 10).

Nature has found several solutions to this problem. In V_κ domains a *cis*-proline in position L8 (sequence numbers according to Kabat et al., 1991) is strongly conserved. In V_λ domains, the kink is one amino acid shorter (structurally, amino acid L7 is missing), and usually contains at least one *trans*-proline (Figure 10). V_H domains, which are similar to V_λ in length, contain glycine residues in positions H8 and/or H10, frequently with positive Phi angles, to make the kink (A. Honegger, unpublished).

It was therefore of interest to investigate which of these solutions to the kink would be selected in a directed evolution experiment. A scFv fragment (specific for a hemagglutinin peptide, PDB file 1ifh) was chosen as a model system, and it contains a kappa light chain. Two small libraries were constructed in which either residues L7, L8 and L9 were randomized, or replaced by only two random amino acids. Thus two libraries were obtained, one of the length of kappa chains, and one of the length of lambda chains at the kink.

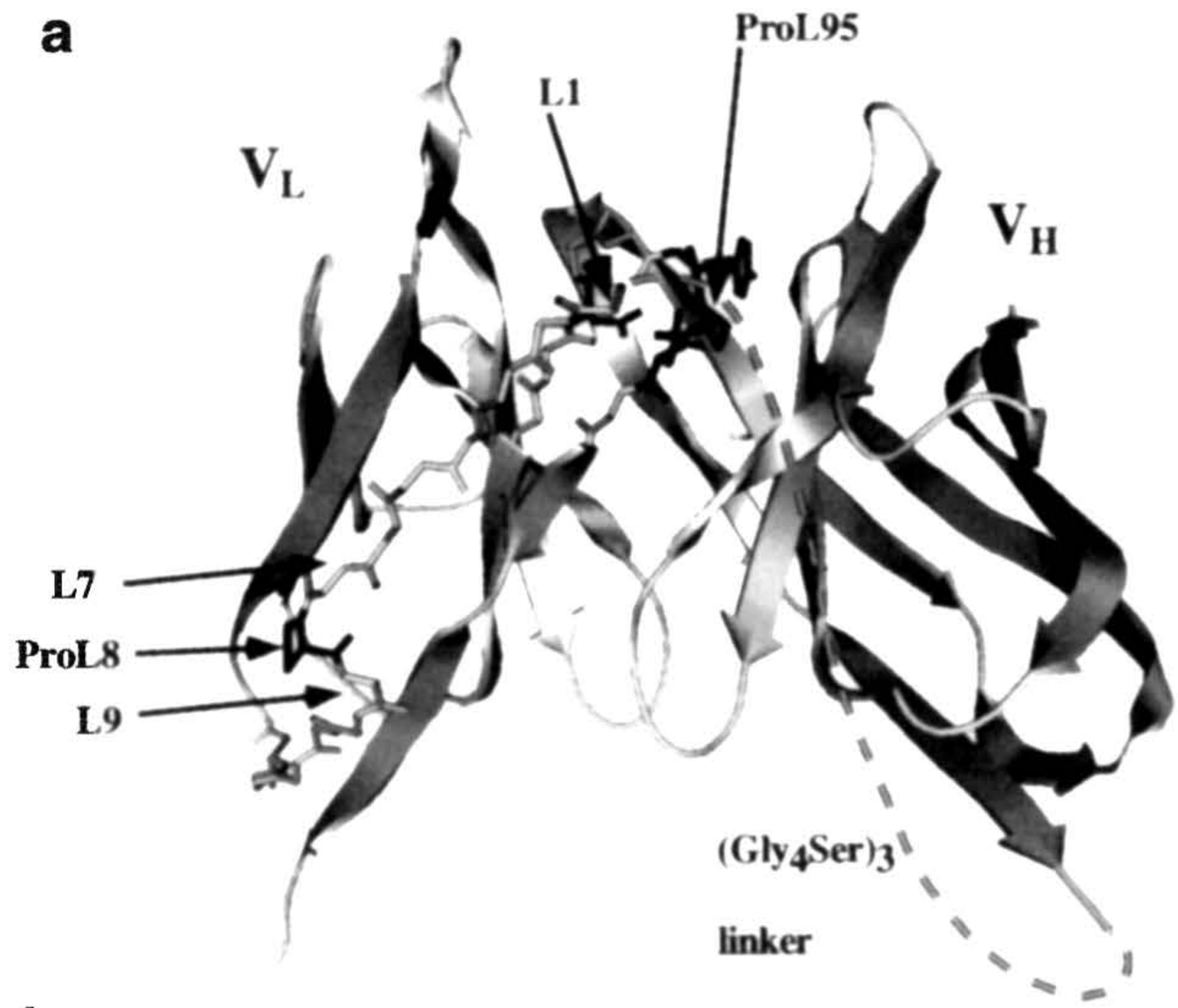
Several rounds of SIP selection were carried out. In the first series of experiments, the "lambda length" variants were kept separate from the kappa length variants. After only one selection round of the "kappa length", 50% of all selected molecules already carried the *cis*-proline typical for these molecules. After a second round, all molecules carried the *cis*-proline. After the third round, the flanking residues began to take shape and very similar preferences were selected as found in the Kabat database of antibody sequences (Fig. 10) (S. Spada, A. Honegger and A. Plückthun, manuscript submitted).

Similarly, in the "lambda length" variants, after only one round, 50% proline were found for either position, which increased to 100% of the molecule having either one of the two of the random positions being proline, with a significant portion having both residues as proline. Interestingly, again these are the variants found by far most frequently in the Kabat database. In a third set of experiments, the "kappa-length" library and the "lambda-length" library were mixed. The kappa sequences were indeed selected, and again a distribution resulted with the most common sequences of kappa domains enriched (S. Spada, A. Honegger and A. Plückthun, manuscript submitted).

Thus, directed molecular evolution reproduces quite faithfully the evolution of the natural molecules, and thus the nature of the selective pressure must have been similar. The molecular parameters why these variants were selected was investigated. The selected variants were found to be more stable against urea denaturation (S. Spada, A. Honegger and A. Plückthun, manuscript submitted). This shows again that complex molecular parameters such as stability can be optimized by molecular evolution.

7. CONCLUSIONS AND PERSPECTIVES

The strategy of using directed molecular evolution is a new tool in the study of proteins. It allows the investigator to concentrate on those mutants which are on the pathway to improved function. As it is quite independent of the availability of working hypotheses, it may possibly uncover new relationships in the protein structure. Clearly, such work does



b

	L7	L8	L9	
kappa				
huma	Ser	*cis*-Pro	Ser,Leu,Ala,Gly	100% *cis*-Pro
mouse	Ser	*cis*-Pro	Ser,Leu,Ala	90% *cis*-Pro
lambda				
huma	—	*trans*-Pro	Ala,*trans*-Pro	
mouse	—	Glu	*trans*-Pro	

Figure 10. (**a**) Location of the cis-prolines L8 and L95 in antibody kappa domains. The antibody is represented according to secondary structure, except in the CDR3 region, and the N-terminal part of the kappa domain, where the backbone is shown. This serves to indicate where the randomized region (L7-L9) is situated with respect to the whole protein. (**b**) The predominant sequences found in kappa and lambda light chains at positions L7, L8 and L9. Lambda chains are one amino acid shorter, and superpositions suggests to make the deletion at position L7.

a

	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
17/9	D I V M T Q S P S S L T V T A G E K V T M S C
library	D I V M T Q X X X S L T V T A G E K V T M S C
	D I V M T Q - X X S L T V T A G E K V T M S C
selected	D I V M T Q S P L S L T V T A G E K V T M S C

b

before selection: X - X - X (no trend)
after 1 round: X - P - X (50% is P)
after 2 rounds: X - P - X (100%is P)
after 3 rounds: S - P - L (50% is sequence)

before selection: X - X (no trend)
after 1 round: P - X (50% is P)
after 2 rounds: P - X (50% is P, some XP and PP)
after 3 rounds: P - X (50% is P, some XP and PP)

before selection: X - X - X and X - X
after 1 round mix: X - P - X (50% is P)
after 2 rounds: S - P - L
S - P - S
S - P - Q

Figure 11. (**a**) The sequence of the starting antibody 17/9 in the N-terminal region of V_L is shown. Two libraries were constructed, containing the same length as in kappa domains (denoted XXX) and one amino acid shorter and thus corresponding to lambda domains (denoted–XX). The most frequently selected molecule has the sequence shown on the bottom. (**b**) The results of the SIP experiment are shown in more detail, as are the predominant features obtained after different numbers of rounds from the "XXX" library and the "XX" library (left), and after combination of the libraries (right). It can be seen that the predominant sequences found in the Kabat library of antibody sequences are remarkably well selected.

not obviate any biophysical investigations whatsoever. On the contrary, since the sequence space is so enormous, the evolutionary approach often still needs to be focused, as a complete sampling is impossible. Thus, structural and evolutionary studies are likely to flourish best when used in combination. The main challenge for the future is to truly "direct" evolution, that is to learn how to apply the pressure exclusively to the molecular quantity of interest. It appears that directed molecular evolution is ultimately a very "rational" approach.

ACKNOWLEDGMENTS

I am indebted to all my coworkers mentioned in the text and references, who have developed the technologies described in this article and have been a constant source for stimulating discussions. Particularly, Dr. Annemarie Honegger has been instrumental in the molecular modeling for the projects described and an invaluable intellectual contributor to all structural issues. I thank Jozef Hanes, Frank Hennecke, Annemarie Honegger, Lutz Jermutus, and Peter Lindner for critically reading the manuscript.

REFERENCES

Baca, M., Scanlan, T. S., Stephenson, R. C. and Wells, J. A. (1997). Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. *Proc. Natl. Acad. Sci. USA* **94**, 10063–10068.

Bai, C. and Elledge, S. J. (1996). Gene identification using the yeast two-hybrid system. *Methods Enzymol.* **273**, 331–347.

Boder, E. T. and Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnol.* **15**, 553–557.

Biocca, S. and Cattaneo, A. (1995). Intracellular immunization: Antibody targeting to subcellular compartments. *Trends Cell Biol.* **5**, 248–252.

Cadwell, R. C. and Joyce, G. F. (1994). Mutagenic PCR. *PCR Methods Appl.* **3**, 136–140.

Cortese, R., Monaci, P., Luzzago, A., Santini, C., Bartoli, F., Cortese, I., Fortugno, P., Galfre, G., Nicosia, A. and Felici, F. (1996). Selection of biologically active peptides by phage display of random peptide libraries. *Curr. Opin. Biotechnol.* **7**, 616–621.

Cull, M. G., Miller, J. F. and Schatz, P. J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc. Natl. Acad. Sci. USA* **89**, 1865–1869.

Dower, W. J. and Cwirla, S. E. (1992) in *Guide to Electroporation and Electrofusion*, Eds.: Chang, D. C., Chassy, B. M., Saunders, J. A. and Sowers, A. E., Academic Press, San Diego, 291–301.

Duenas, M. and Borrebaeck, C. A. (1994). Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. *Bio/Technol.* **12**, 999–1002.

Dunn, I. S. (1996). Phage display of proteins. *Curr. Opin. Biotechnol.* **7**, 547–553.

Fersht, A. R. (1995). Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84.

Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9.

Fersht, A. R. and Wells, T. N. (1991). Linear free energy relationships in enzyme binding interactions studied by protein engineering. *Protein Eng.* **4**, 229–231.

Freedman, R. B., Hirst, T. R. and Tuite, M. F. (1994). Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem. Sci.* **19**, 331–336.

Frisch, C., Kolmar, H. and Fritz, H. J. (1994). A soluble immunoglobulin variable domain without a disulfide bridge: Construction, accumulation in the cytoplasm of E. coli, purification and physicochemical characterization. *Biol. Chem. Hoppe-Seyler* **375**, 353–356.

Frisch, C., Kolmar, H., Schmidt, A., Kleemann, G., Reinhardt, A., Pohl, E., Usón, I., Schneider, T. R. and Fritz H. J. (1996). Contribution of the intramolecular disulfide bridge to the folding stability of REI(v), the variable domain of a human immunoglobulin kappa light chain. *Fold. Des.* **1**, 431–440.

Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. and Curtiss, R,III (1997). *Nature Biotechnol.* **15,** 29–34.

Gilbert, H. F. (1990). Molecular and cellular aspects of thiol-disulfide exchange. *Adv. Enzymol.* **63**, 69–172.

Glockshuber, R., Schmidt, T. and Plückthun, A. (1992). The disulfide bonds in antibody variable domains: effects on stability, folding *in vitro*, and functional expression in *Escherichia coli. Biochemistry* **31**, 1270–1279.

Goto, Y. and Hamaguchi, K. (1979). The role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain. *J. Biochem.* **86**, 1433–1441.

Gramatikoff, K., Georgiev, O. and Schaffner, W. (1994). Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. *Nucleic Acids Res.* **22**, 5761–5762.

Hall, A. and Knowles, J. R. (1976). Directed selective pressure on a beta-lactamase to analyse molecular changes involved in development of enzyme function. *Nature* **264**, 803–804.

Hanes, J. and Plückthun, A. (1997). *In vitro* selection and evolution of functional proteins using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942.

Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, 121–126.

Huston, J. S., McCartney, J., Tai, M. S., Mottola-Hartshorn, C., Jin, D., Warren, F., Keck, P. and Oppermann, H. (1993). Medical applications of single-chain antibodies. *Intern. Rev. Immunol.* **10**, 195–217.

Jin, L., Fendly, B. M. and Wells, J. A. (1992). High resolution functional analysis of antibody-antigen interactions. *J. Mol. Biol.* **226**, 851–865.

Jung, S. and Plückthun, A. (1997). Improving *in vivo* folding and stability of a single-chain Fv antibody fragment by loop grafting. *Prot. Eng.* **10**, 959–966.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. and Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th Ed., National Institutes of Health, Bethesda, MD.

Kayushin, A. L., Korosteleva, M. D., Miroshnikov, A. I., Kosch, W., Zubov, D. and Piel, N. (1996). A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries. *Nucleic Acids Res.* **24**, 3748–3755.

Keiler, K. C., Waller, P. R. and Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993.

Keohavong, P. and Thilly, W. G. (1989). Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* **86**, 9253–9257.

Knappik, A. and Plückthun, A. (1995). Engineered turns of an antibody improve its *in vivo* folding. *Protein Eng.* **8**, 81–89.

Krebber, C., Moroney, S., Plückthun, A. and Schneider, C. (1993). European Patent Application EP 93102484.

Krebber, C., Spada, S., Desplancq, D. and Plückthun, A. (1995). Coselection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Lett.* **377**, 227–231.

Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. and Plückthun, A. (1997). Selectively infective phage (SIP): a mechanistic dissection of a novel method to select for protein-ligand interactions. *J. Mol. Biol.* **268**, 619–630.

Lerner, R. A., Benkovic, S. J. and Schultz, P. G. (1991). At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **252**, 659–667.

Low, N. M., Holliger, P. H. and Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.* **260**, 359–368.

Lubkowski, J., Hennecke, F., Plückthun, A. and Wlodawer A. (1998). The structural basis of phage display: The crystal structure of the N-terminal domains of g3p at 1.46 Å resolution. *Nature Structural Biology*, **5**, 140–147.

Mattheakis, L. C., Bhatt, R. R. and Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* **91,** 9022–9026.

Moore, J. C., Jin, H. M., Kuchner, O. and Arnold, F. H. (1997). Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* **272**, 336–347.

Nieba, L., Honegger, A., Krebber C. and Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10**, 435–444.

Padlan, E. A. (1996). X-ray crystallography of antibodies. *Adv. Protein Chem.* **49**, 57–133.

Phizicky, E. M. and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123.

Proba, K., Wörn, A., Honegger, A. and Plückthun, A. (1998). Antibody fragments without disulfide bonds, made by molecular evolution. *J. Mol. Biol.* **275**, 245–253.

Richardson, J. H. and Marasco, W. A. (1995). Intracellular antibodies: development and therapeutic potential. *Trends Biotechnol.* **13,** 306–310.

Rudikoff, S. and Pumphrey, J. G. (1986). Functional antibody lacking a variable-region disulfide bridge. *Proc. Natl. Acad. Sci. USA* **83,** 7875–7878.

Ryabova, L., Desplancq, D., Spirin, S. and Plückthun, A. (1997). Making antibodies *in vitro*: the effect of molecular chaperones and disulfide isomerase on the functional expression of single-chain fragments by *in vitro* translation. *Nat. Biotechnol.* **15**, 79–84.

Scott, J. K. and Smith, G. P. (1990). Searching for peptide ligands with an epitope library. *Science* **249**, 386–390.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317.

Spada, S. and Plückthun, A. (1997). Selectively infective phage (SIP) technology: a novel method for the *in vivo* selection of interacting protein-ligand pairs. *Nature Med.* **6**, 694–696.

Spada, S., Krebber, C. and Plückthun, A. (1997). Selectively infective phages. *Biol. Chem.* **378**, 445–456.

Stemmer, W. P. (1994a). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370,** 389–391.

Stemmer, W. P. (1994b). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751.

Usón, I., Bes, M. T., Sheldrick, G. M., Schneider, T. R., Hartsch, T. and Fritz, H. J. (1997). X-ray crystallography reveals stringent conservation of protein fold after removal of the only disulfide bridge from a stabilized immunoglobulin variable domain. *Fold. Des.* **2,** 357–361.

Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G. and Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucl. Acids Res.* **22**, 5600–5607.

Wagner, S. D. and Neuberger, M. S. (1996). Somatic hypermutation of immunoglobulin genes. *Annu. Rev. Immunol.* **14**, 441–457.

Warbrick, E. (1997). Two's company, three's crowd: the yeast two hybrid system for mapping molecular interactions. *Structure* **5,** 13–17.

Ward, W. H., Timms, D. and Fersht, A. R. (1990). Protein engineering and the study of structure-function relationships in receptors. *Trends Pharmacol. Sci.* **11**, 280–284.

Wells, J. A., Cunningham, B. C., Fuh, G., Lowman, H. B., Bass, S. H., Mulkerrin, M. G., Ultsch, M. and deVos, A. M. (1993). The molecular basis for growth hormone-receptor interactions. *Recent Prog. Horm. Res.* **48**, 253–275.

Winter, G., Griffiths, A. D., Hawkins, R. E. and Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455.